

# MScFE 632 Machine Learning in Finance

## Group Work Project # 1

[See grading rubric here.](#)

### Scenario

Suppose you are a team of alpha quants on the strategies desk. As your first responsibility, you will be tasked with modeling using cutting-edge methodologies for finding trading strategies for investment managers. This is a very important—and highly visible—role. Perform it correctly and your group's expertise will be known as Team Alpha, which is attached to a robust revenue stream. Perform it incorrectly and...well, let's not do that. There are many tools you learned in the first three weeks of machine learning. These include:

- Category 1:
  - LASSO regression
  - RIDGE regression
  - Elastic nets
- Category 2:
  - k-means clustering
  - Hierarchical clustering
- Category 3:
  - Principal components
- Category 4:
  - Classification trees
  - Regression trees

### Tasks

Your group will create a **marketing handbook** that provides thoughtful guidelines for addressing some of these challenges. You will pick three of the categories above. For each category, the group only needs to select one of the topics (e.g., for Category 4, the group would select either Classification trees OR Regression trees, not both). Then you will write a report addressing each of the following areas. Any code derived from this analysis must be in a Jupyter notebook.

- **Advantages:** The benefits of using this methodology
- **Basics:** Definition and classification (e.g., this is a linear regression that...)
- **Computation:** A Jupyter notebook that illustrates the method
- **Disadvantages:** Difficulties or known issues with this methodology
- **Equations:** Equations that summarize how the model works
- **Features:** Features of the model (e.g., works well with missing values)
- **Guide:** List of inputs and outputs
- **Hyperparameters:** List of hyperparameters that need tuning
- **Illustration:** Visuals (figures, flowcharts, graphs) that show HOW the model works; can cite existing references
- **Journal:** One reference to a journal article (no Wikipedia page or blog) that illustrates the use of that technique as it is applied to finance. Do not include a required reading from this course.
- **Keywords:** Create “tags” that identify this model

## Step 1

**As a group**, all three members select the three categories and work together on writing the basics and keywords.

**Note: If there are only 2 students in the group**, then the pair decides on only two items. Each member does one individually.

## Step 2

**Individually**, each student is responsible for one challenge by writing its:

- **Advantages:** The benefits of using this methodology
- **Computation:** A Jupyter notebook that illustrates the method
- **Disadvantages:** Difficulties or known issues with this methodology
- **Equations:** Equations that summarize how the model works
- **Features:** Features of the model (e.g., works well with missing values)
- **Guide:** List of inputs and outputs
- **Hyperparameters:** List of hyperparameters that need tuning
- **Illustration:** Visuals (figures, flowcharts, graphs) that show HOW the model works; can cite existing references

- **Journal:** One reference to a journal article (no Wikipedia page or blog) that illustrates the use of that technique as it is applied to finance. Do not include a required reading from this course.
- **Keywords:** Create “tags” that identify this model.

For example:

- Team Member A works on Model 1
- Team Member B works on Model 2
- Team Member C works on Model 3

### Step 3

As a group, the team members work together on the section called “**Technical Section**” where they describe how models in general have their hyperparameters tuned. This section should include examples of hyperparameters used from the individual reports.

### Step 4

As a group, the team members work together on a section called “**Marketing Alpha**,” which integrates the advantages and features from the individual reports to show how ML techniques can do well. The result is 1–2 pages.

### Step 5

As a group, the team members work together on a section called “**Learn More**,” which combines the references to the journal articles from the individual reports and other ML websites that emphasize the strengths of ML algorithms (as opposed to websites that teach how to do it or provide code samples). Be sure the references are in proper MLA format.

## Submission Requirements and Format

One team member submits the following on behalf of the entire group:

- 1) A **zipped folder** including:
  - a) A single executable Jupyter notebook\* that addresses all the computations
  - b) A duplicate version of the Jupyter notebook code and output in PDF or HTML format.

- i) In order to include the output of the code, you must RUN the code before downloading the PDF.
- 2) **One PDF document** with all sections EXCEPT ANY CODE. This PDF should just contain text, formulas, and graphs, but no Python or other code. In fact, this document should contain all the sections EXCEPT any parts or sections that have code. Please be sure that code only appears in the two files above.
  - a) Use the available Report Template and fill out the required information on the first page.

***\*Use Google Colab or GitHub to collaborate** in completing the executable Python program.*

*The PDF file with the answers to the questions must be uploaded **separately** from the zipped folder that includes any other types of files. This allows Turnitin to generate a similarity report.*

## Rubric

Your instructor will evaluate your group submission for GWP1 using the following rubric:

| Quantitative Analysis<br>(Open-Ended Questions)  | Technical and Non-Technical Reports  | Writing and Formatting   |
|--|--|--|
| 40 Points  | 30 Points  | 20 Points  |
| <p>The group is able to apply results, formulas, and their knowledge of theory to real-life finance scenarios by doing the following:</p> <ul style="list-style-type: none"> <li>• Providing all the necessary information to support their arguments.</li> <li>• Presenting arguments that reflect group discussion and research.</li> <li>• Using authoritative references to support a position and provide updated information.</li> <li>• Concluding with practical takeaways for more insightful financial decision-making.</li> </ul> | <p>Technical Reports contain 3 parts:</p> <ol style="list-style-type: none"> <li>1) <b>code</b> for each question (be sure to explicitly state the question number),</li> <li>2) the corresponding output of that <b>code</b>, and</li> <li>3) <b>interpretations and/or recommended courses of action that reasonably follow from those results.</b></li> </ol> <p><b>Note:</b> Technical reports will include the technicalities of models, such as names, methods of estimation, parameter values, etc., and exclude generalities about the work done. It should NOT include names of Python code that were used.</p> | <p>A submission that looks professional should:</p> <ul style="list-style-type: none"> <li>• Include the axes, labels, and scales in graphs.</li> <li>• Be free of significant grammatical errors or typos.</li> <li>• Be an organized, well-structured, and easy-to-read document.</li> <li>• Include proper citations and a bibliography in MLA format.</li> </ul> |
|  | <p>Non-Technical Reports contain 3 parts:</p> <ol style="list-style-type: none"> <li>1) clear explanation of results;</li> <li>2) the recommended course of action that follows; and</li> <li>3) the identification of factors that impact each portfolio.</li> </ol> <p><b>Note:</b> AVOID all references to model names, algorithms, and unnecessary details. Instead, focus on the investment decision.</p>   |  |